DOCUMENT RESUME

ED 397 096 TM 025 138

AUTHOR	Salvucci, Sameena; And Others
TITLE	Design Effects and Generalized Variance Functions for the 1990-91 Schools and Staffing Survey (SASS). Volume II. Technical Report.
INSTITUTION	Synectics for Management Decision, Inc., Arlington, VA.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	ISBN-0-16-048504-5; NCES-95-342-II
PUB DATE	Feb 95
CONTRACT	RN-91-0600.01
NOTE	53p.; For Volume I, see TM 025 137.
AVAILABLE FROM	U.S. Government Printing Office, Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328.
PUB TYPE	Statistical Data (110) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Administrators; *Computation; Elementary Secondary Education; *Error of Measurement; Estimation (Mathematics); Institutional Characteristics; *National Surveys; *Research Design; Sample Size; *Sampling; Student Characteristics; Teacher Characteristics; Teacher Supply and Demand
IDENTIFIERS	*Generalized Variance Functions; *Schools and Staffing Survey (NCES)

ABSTRACT

This technical report provides the results of a study on the calculation and use of generalized variance functions (GVFs) and design effects for the 1990-91 Schools and Staffing Survey (SASS). The SASS is a periodic integrated system of sample surveys conducted by the National Center for Education Statistics (NCES) that produces sampling variances different from those produced by simple random sampling with fixed sample size. An introduction provides a conceptual and contextual discussion, addressing the reasoning for using general analytical techniques to calculate standard errors for complex survey data. The second section describes groups of survey statistics users might want to select. In Section 3, the procedure and computational formulas for calculating the design effects for the most common types of estimates (totals, means, and proportions) are presented. Section 4 then gives the technical details of GVF fitting. The design effect and GVF results obtained from the 1990-92 SASS are presented in Section 5. Section 6 discusses possible next steps in calculating standard errors for NCES surveys. (Contains 8 tables, 12 figures, and 25 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 397 096

TM

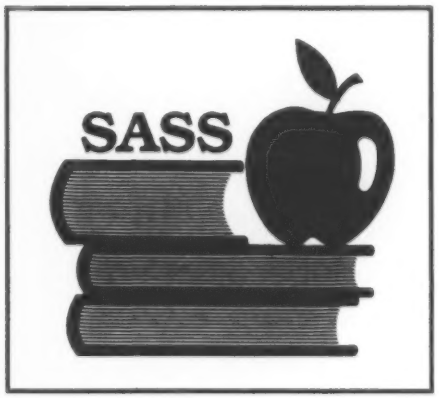
NATIONAL CENTER FOR EDUCATION STATISTICS

Design Effects and Generalized Variance Functions for the 1990–91 Schools and Staffing Survey (SASS)

Volume II

Technical Report

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.



U.S. Department of Education
Office of Educational Research and Improvement NCES 95-342-II

BEST COPY AVAILABLE

TM025138

NATIONAL CENTER FOR EDUCATION STATISTICS

Design Effects and Generalized Variance Functions for the 1990–91 Schools and Staffing Survey (SASS)

Volume II

Technical Report

Sameena Salvucci
Albert Holt
Ramal Moonesingle
Synectics for Management Decisions, Inc.

Steven Kaufman, Project Officer
National Center for Education Statistics

U.S. Department of Education
Office of Educational Research and Improvement NCES 95-342-II

U.S. Department of Education
Richard W. Riley
Secretary

Office of Educational Research and Improvement
Sharon P. Robinson
Assistant Secretary

National Center for Education Statistics
Jeanne E. Griffith
Acting Commissioner

National Center for Education Statistics

The purpose of the Center is to collect and report "statistics and information showing the condition and progress of education in the United States and other nations in order to promote and accelerate the improvement of American education."—Section 402(b) of the National Education Statistics Act of 1994 (20 U.S.C. 9001).

February 1995

Contact:
Steven Kaufman
(202) 219–1337

For sale by the U.S. Government Printing Office
Superintendent of Documents, Mail Stop: SSOP, Washington, DC 20402-9328
ISBN 0-16-048504-5

ACKNOWLEDGMENTS

This series of reports was prepared by Synectics for Management Decisions, Inc., a contractor to the National Center for Education Statistics, under Contract No. RN-91-0600.01.

The authors wish to thank all of those who contributed to the production of this report. Among Synectics staff, special mention goes to Fan Zhang, Mehrdad Saba, Michael Chang, and Nagarthun Movva who provided very capable programming support; Mayra Walker and Steven Fink who prepared the tables; and Elizabeth Walter, who carefully edited the report.

Steven Kaufman and Dan Kasprzyk, of the Elementary and Secondary Education Statistics Division, reviewed the manuscript through several drafts. We would also like to acknowledge the helpful comments of the following technical reviewers: Marilyn McMillen, Steve Broughman, Michael P. Cohen, Mary Rollefson, Randall J. Parmer, and Betty Jackson.

CONTENTS

Section			Page
	Preface		vii
1	Introduction		1
	1.1	1990-91 SASS	3
	1.2	Sample Design	4
	1.3	Accuracy of Estimates	6
2	Groups of Survey Statistics		9
3	Design Effect Methodology		13
	3.1	Design Effect for Totals	14
	3.2	Design Effect for Means	15
	3.3	Design Effect for Proportions	16
	3.4	Average Design Effects	17
4	GVF Methodology		19
	4.1	GVF Models	19
	4.2	GVF Estimation for 1990-91 SASS	24
5	Results and Conclusions		31
	5.1	Average Design Effects	31
	5.2	"Best" GVF Models	33
	5.3	GVF Validation	35
	5.4	Reduction of GVF Tables	39
	5.5	User's Manual	42
6	Next Steps		43
	References		45

PREFACE

This technical report provides the results of a study on the calculation and use of generalized variance functions (GVFs) and design effects for the 1990-91 Schools and Staffing Survey (SASS). It is Volume II of a two-volume publication that is part of the Technical Report Series published by the National Center for Education Statistics (NCES). Volume I, the User's Manual, is written to illustrate the application procedures for calculating standard errors using the design effects and generalized variance functions for the 1990-91 SASS as produced by this study.

The structure of this volume reflects the belief that different readers may come to it with dissimilar goals.

Section 1: Introduction

Some readers will need a conceptual and contextual discussion, addressing the reasoning for using general analytical techniques to calculate standard errors for complex survey data. This section addresses such questions such as "Why are design effects and generalized variance functions useful in SASS?" and "What is the sample design for the 1990-91 SASS?"

Section 2: Groups of Survey Statistics

We tried to anticipate the range of interests users might have in selecting the various combination of statistics for their analyses. Descriptions of these groups of statistics are included. An example of a statistic that might be of interest is the total number of students enrolled in first grade from the School Survey. If a user was interested in calculating the standard error of this estimate they would use the results of this study for the group of statistics labeled "Student Totals" from the School Survey.

Section 3: Design Effect Methodology

Some readers will find a measure of the efficiency of the SASS design of interest. This section provides the procedure and computational formulas for calculating the design effects for the most common types of estimates: totals, means, and proportions.

Section 4: GVF Methodology

The technical details of the GVF fitting may be the main interest of some readers. This section provides formulas for five types of common GVFs and the results of three different fitting methodologies. Both the properties of GVFs addressed in this work and the results of GVF fitting have limitations which require discussion.

Section 5: Results and Conclusions

Many may wish to go directly to the design effect and GVF results obtained for the 1990-91 SASS and contrast these results with earlier GVF information using SASS 1987-88 data.

The average design effect and GVF tables produced by this study are provided in the appendices of Volume I, the User's Manual.

Section 6: Next Steps

The report concludes with brief remarks on possible next steps.

1. Introduction

The Schools and Staffing Survey (SASS) is a periodic, integrated system of sample surveys conducted by the National Center for Education Statistics (NCES) of the U.S. Department of Education. The complex sample design of SASS produces sampling variances different from those produced by *simple random sampling* (srs) with fixed sample size. This is so for a number of reasons. There are gains in precision from stratification by geography, type of school, size of school, and so on. These gains, however, are counterbalanced by the effects of clustering of students and teachers within sampled schools. Weighting can be conducted to determine the contribution of sample units to the population estimates. However, the weights themselves are subject to sampling variability which may make nonlinear the statistics which are linear with simple random sampling. The calculation of variance estimates for SASS statistics are, therefore, more complex than the simple random sample variance estimation algorithms and computationally more expensive. Using the simple random sample methods for SASS complex samples almost always underestimates the true sampling variances and makes differences in the estimates appear to be significant when they are not. Unfortunately, general use statistical packages such as SAS, SPSS, etc., only calculate sampling variances based on simple random sample and are thus not appropriate for estimating variances for SASS.

SASS provides data on public and private schools, public school districts, teachers, and administrators, and is used by educators, researchers, and policy makers. The SASS data sets contain approximately 1,500 variables. In addition, statistics such as totals, averages, proportions, differences, and many others can also be estimated. Although calculation and publication of a separate sampling error (sampling variance or its square root, the standard error) for each estimate might be possible with today's computing power, there are practical reasons, as well as methodological motivation, for more general analytical techniques desirable to produce stable and precise sampling variance estimates. These reasons (Wolter 1985 and Hanson 1978) are mentioned briefly here with reference to SASS:

- Presentation of individual sampling errors would double the number of tables in a report, increasing computing, printing, and associated personnel costs

- Only computer-readable SASS public-use files are available to compute estimates that do not appear in publications.

- Each SASS public-use file includes a set of 48 variables for replicate weights. These replicate weights were designed to produce variances using the balanced half-sample replication technique. However, these replicate weights can be utilized only by users who have half-sample replication software available.

- By averaging over time or generalizing in some way, more stable sampling error estimates can be produced.

- In repeated execution of surveys of the same population and with the same types of variables, it might be possible to use parameter estimates from earlier applications for developing generalized variance models.

- An accurate generalized variance model may also be of great value in designing similar surveys in the future.

These considerations have led to the use of simple mathematical models as a means of approximating sampling errors. These mathematical models, known as *generalized variance functions* (GVFs), relate the variance or *relative variance* of a survey estimate to the mean (expectation) of the estimate, where the relative variance is the variance divided by the square of the mean. Section 4 will discuss in detail the GVF methodology. Two major surveys in the United States that have used GVFs are the Current Population Survey (CPS) and the National Health Interview Survey (NHIS).

There are several mathematical models available as possible candidates to be a GVF for groups of statistics in a survey. The degree of fit for each of the models must be examined and the model with the best fit would be selected. The usual practice in large scale and complex sample design surveys like SASS is to use a set of sampling errors (variances, or standard errors), estimated directly by replication method, to estimate the parameters (usually by the least squares method) of a well-chosen GVF. The estimated parameter values are published or used to generate tables so that users can approximate sampling errors for a variety of statistics simply by evaluating the model at the survey estimates.

Valliant (1987) shows that in some data settings generalized variance functions perform as well or better than direct variance estimators in terms of bias, precision, and confidence interval construction. The performance of the GVFs generally depends on the critical issue of grouping of a set of survey estimates for GVF modeling, and the type of GVF model chosen including the method of estimating the parameters of the GVF model. However, a cautionary note is that there are likely to be survey variables (e.g., rare characteristics) whose GVF model differs considerably from that of most variables and for which GVF will give poor results. Section 3.4 of volume I of this publication provides a list of specific types of variables in SASS for which GVF may be inappropriate.

As known, the SASS complex sample differs from the simple random sample. The calculation of variance estimates for SASS statistics are more complex than the simple random sample variance estimation. The impact of the complex design on the reliability of a sample estimate, in comparison to the alternative simple random sampling, is often measured by the *design effect* (Deff), which will be discussed in detail in section 3 of this volume. The notions of design effect and average design effect have helped develop the generalized variance

functions. It is useful to calculate the average design effect for a group of survey estimates used to develop the GVF model. Design effect also provides an alternative way to approximate the sampling variance estimates.

This report summarizes the results of an empirical study on the calculation and properties of generalized variance functions and design effects as applied to the 1990-91 SASS estimators of totals, averages, and proportions. The following sections provide an overview of the 1990-91 SASS sample design and estimation (Kaufman and Huang 1993).

1.1 1990-91 SASS

The data were obtained in the second cycle of the Schools and Staffing Survey (SASS) conducted by the National Center for Education Statistics (NCES) in 1990-91. SASS provides data on public and private schools, public school districts, teachers, and administrators, and is used by educators, researchers, and policy makers. The survey includes several types of respondents: school district personnel, public school principals, private school principals, public school teachers, and private school teachers. The 1990-91 SASS is a set of four interrelated national surveys.

The following elements make up the 1990-91 SASS:

(1) The Teacher Demand and Shortage (TDS) Survey targeted public school district personnel who provided information about their district's student enrollment, number of teachers, position vacancies, new hires, teacher salaries and incentives, and hiring and retirement policies.

(2) The School Administrator Survey collected background information from principals on their education, experience, and compensation and also asked about their perceptions of the school environment and the importance they place on various educational goals.

(3) The School Survey included information on student characteristics, staffing patterns, student-teacher ratios, types of programs and services offered, length of school day and school year, graduation and college application rates, and teacher turnover rates. The 1990-91 private school questionnaire incorporated questions on aggregate demand for both new and continuing teachers.

(4) The Teacher Survey collected information on public and private school teachers' demographic characteristics, education, qualifications, income sources, working conditions, plans for the future, and perceptions of the school environment and the teaching profession.

1.2 Sample Design

The target populations for the 1990-91 SASS surveys included U.S. elementary and secondary public and private schools with students in any of grades 1-12, principals and classroom teachers in those schools, and local education agencies (LEAs) that employed elementary and/or secondary level teachers. In the private sector, since there is no counterpart to the LEAs, information on teacher demand and shortages was collected directly from individual schools. The SASS sample was designed to produce (1) national estimates for public and private schools, (2) state estimates for public schools, (3) state/elementary, state/secondary, and national combined public school estimates, and (4) detailed association estimates and grade level estimates for private schools.

These are the three primary steps in the sample selection process followed during the 1990-91 SASS:

(1) A sample of schools was selected. The same sample was used for the School Administrator Survey.

(2) Each LEA that administered one or more of the sample schools in the public sector became part of the sample for the Teacher Demand and Shortage Survey. For the sample of private schools, the questions for the Teacher Demand and Shortage Survey were included in the questionnaire for the School Survey.

(3) For each sample school, a list of teachers was obtained from which a sample was selected for inclusion in the Teacher Survey.

Details pertaining to the frame, stratification, sorting, and sample selection for each of the four surveys of SASS are described in the subsections below (Kaufman and Huang 1993).

1.2.1 School Survey

The School Survey had two components: private schools and public schools. The primary frame for the public school sample was the 1988-89 Common Core of Data (CCD) file. The CCD survey includes an annual census of public schools, obtained from the states, with information on school characteristics and size. A supplemental frame was obtained from the Bureau of Indian Affairs, containing a list of tribal schools and schools operated by that agency. The school sample was stratified, with the allocation of sample schools among the strata designed to provide estimates for several analytical domains. Within each stratum, the schools in the frame were further sorted on several geographic and other characteristics. A specified number of schools were selected from each stratum with probability proportionate to the square root of the number of teachers as reported on the CCD file. The target sample size of public schools was 9,687.

A dual frame approach was used to select the samples of private schools. A list frame was the primary private school frame, and an area frame was used to find schools missing from the list frame, thereby compensating for the coverage problems of the list frame. To supplement the list frame, an area sample consisting of 123 primary sampling units (PSUs) was selected. The target sample size of private schools was 3,270, with 2,670 allocated to the list sample and 600 to the area sample. The list sample was allocated to 216 strata defined by association group, school level (elementary, secondary, combined), and census region (northeast, midwest, south, west). There were 18 association groups; for example, Catholic, National Society of Hebrew Day Schools, and National Association of Independent Schools. Within each stratum, schools were sorted by state and other variables within state. The area sample was allocated to strata defined by 123 PSUs and school level (elementary, secondary, combined). Within each stratum, schools were sorted by affiliation (Catholic, other religious, and nonsectarian), 1989 PSS enrollment, and school name. For both the list sample and the area sample, schools were systematically selected from each stratum with probability proportionate to the square root of the number of teachers as reported in the 1989-90 PSS. Any school with a measure of size larger than the sampling interval was excluded from the probability sampling operation and included in the sample with certainty.

1.2.2 School Administrator Survey

For the School Administrator Survey the target population consisted of the administrators of all public and private schools eligible for inclusion in the School Survey. Once the sample of schools was selected, no additional sampling was needed to select the sample of school administrators. Thus, the target sample size was the same as for the School Survey (n=12,957). Some of these schools did not have administrators, in which case the school was asked to return the questionnaire, but, with few exceptions, there was a one-to-one correspondence between the SASS samples of schools and school administrators.

1.2.3 Teacher Demand and Shortage Survey

The Teacher Demand and Shortage (TDS) Survey has two components: public schools and private schools.

For the public school sector, the target population consisted of all U.S. public school districts. These public school districts, often called local education agencies (LEAs), are government agencies administratively responsible for providing public elementary and/or secondary education. LEAs associated with the selected schools in the school sample received a TDS questionnaire. An additional sample of districts not associated with schools was selected and also received the TDS questionnaire. The target sample size was 5,424.

For the private school sector, the target population consisted of all U.S. private schools. Thus, the target sample size was the same as the private school sample of 3,270. The school questionnaire for the selected private schools included TDS questions for the school.

1.2.4 Teacher Survey

The target population for the Teacher Survey consisted of full-time and part-time teachers whose primary assignment was teaching in kindergarten through grade 12. Data were collected from a sample of classroom teachers in each of the public and private schools that was included in the sample for the School Survey: the selected schools were asked to provide teacher lists for their schools and then those lists were used to select 56,051 public and 9,166 private school teachers. The survey designs for the public and private sectors were very similar. Within each selected school, teachers were stratified into one of five types in hierarchical order, as 1) Asian or Pacific Islander, 2) American Indian, Aleut, or Eskimo, 3) Bilingual/ESL (English as a Second Language), 4) New (less than three years teaching experience), or 5) Experienced (three or more years of teaching experience). Within each stratum, teachers were selected systematically with equal probability.

1.3 Accuracy of Estimates

The final sample of respondents for each of the four 1990-91 SASS surveys provided measures of characteristics of schools, students, teachers, and administrators. Estimates of means, totals, and proportions obtained for these items involve weights that reflect adjustments for nonresponse and poststratification.

The SASS design described above will produce variances different from the variances produced by simple random sampling (srs) with fixed sample size. This is so for a number of reasons. There are gains in precision from stratification by geography, type of school, and size of school. These gains, however, are counterbalanced by the effects of clustering. The weights themselves are subject to sampling variability which makes nonlinear the statistics which are linear with simple random sampling. The estimators of sampling variances for SASS statistics are, therefore, more complex than the simple random sample estimation algorithms and computationally more expensive.

A class of techniques, called *replication* methods, provides a general approach of estimating variances for the types of sample designs and weighting procedures usually encountered in complex sample surveys. Essentially, the idea behind the replication approach is to repeatedly select portions of the sample to calculate the estimate of interest and then use the variation among these quantities to estimate the variance of the full sample statistics.

These subsamples are called *replicates*, and the estimates calculated from these replicates are called *replicate estimates*.

The *balanced half-sample replication* (also called *balanced repeated replication*, abbreviated as BRR) method has been used to estimate the sampling errors associated with estimates for all of the 1990-91 SASS surveys. In the BRR methodology, within each stratum, sampled schools are paired by the order they were selected. One school from each pair is placed into each replicate. Each replicate includes approximately half the total sample, hence the name half-sample replication. The choice of when to place a school from a pair into a replicate is done in a balanced manner to reduce the variability of the variance estimates. See Kaufman and Huang (1993) for more information on how SASS units are placed into balanced half-sample replicates. Given the replicate weights, the statistic of interest, such as the number of kindergarten teachers from the School Survey, is estimated from the full sample and from each replicate. The mean square error of the replicate estimates around the full sample estimate provides an estimate of the variance of the statistic. By formula, the BRR variance estimate is expressed as follows:

$$V(\hat{X}) = \frac{1}{G}\sum_{k=1}^{G}(\hat{X}_k - \hat{X})^2$$

where $\hat{X}$ is the full sample estimate of X, the statistic of interest, $\hat{X}_k$ is the k-th replicate estimate of X, and G is the number of the replicates.

SASS uses 48 replicates for variance estimation. Optimally, each replicate corresponds to one degree of freedom in a t-test of significance. A minimum of 30 replicates is required for the t-test to be approximated by a z-test. The BRR replicates are not independent. However, if the stratum variances are all the same, then the degrees of freedom will equal the number of strata (which is almost the same as the number of replicates), the dependence of the replicates notwithstanding. To the extent that the stratum variances vary, the degree of freedom are reduced. Forty-eight replicates give a reasonable degree of freedom cushion for the validity of the z-test approximation.

NCES has prepared public use data files for the 1990-91 SASS which include the set of 48 weighted replicates. However, these replicates can be utilized only by users who have software available to perform the balanced half-sample replication estimation. One instance of such software is a SAS (Statistical Analysis System) user-written procedure called Proc WESVAR (Westat 1993) which computes basic survey estimates and their associated sampling errors for user-specified characteristics. For examples of other software that support BRR, see Wolter (1985).

2. Groups of Survey Statistics

NCES publishes statistics for many characteristics and some standard subpopulations. Based on these publications, and in anticipation of various combinations of results (e.g., totals, averages, proportions) which may be of interest to users, table 2.1 below lists the groups of statistics examined for each of the SASS surveys during the GVF modeling procedure. From a substantive point of view, the groupings will often be successful when the statistics refer to (1) the same demographic or economic characteristic, (2) the same race-ethnicity group, and (3) the same level of geography (Wolter 1985). Table 2.2 describes the relevant subpopulations for each group of statistics in the four surveys, and table 2.3 provides definitions of each subpopulation included in this study. The grouping described in this section then gives the frame of the GVF, as well as design effect, tables produced by this study, as provided in Appendices II and III, Volume I, User's Manual, of this publication.

Table 2.1 -- Groups of statistics in 1990-91 SASS GVF study

Survey	Group of statistics
School	Student Totals (e.g., number of students enrolled in 1st grade) Teacher Totals (e.g., number of full-time K-12 teachers) Student Averages (e.g., average number of ungraded students) Teacher Averages (e.g., average number of Hispanic K-12 teachers) School Proportions (e.g., proportion of schools offering kindergarten)
School Administrator	Administrator Totals (e.g., number of administrators with master's degrees) Administrator Averages (e.g., average age of administrators) Administrator Proportions (e.g., proportion of male administrators)
Teacher Demand and Shortage (Private)	TDS Totals (e.g., number of full-time equivalent teachers with state certification) TDS Proportions (e.g., proportion of districts with retraining offered teachers: special education)
Teacher Demand and Shortage (Public)	Student Totals (e.g., number of ungraded students) Teacher Totals (e.g., number of full-time equivalent grade 1-6 teachers) Student Averages (e.g., average number of prekindergarten students) Teacher Averages (e.g., average number of postsecondary teachers) Proportions (e.g., proportion of math teachers offered retraining)
Teacher	Teacher Totals (e.g., number of male teachers) Teacher Averages (e.g., average number of years as a part-time teacher) Teacher Proportions (e.g., proportion of married teachers)

SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91.

Table 2.2 -- Relevant subpopulations for groups of statistics in 1990-91 SASS

Survey	Subpopulation for each group of statistics
School	Sector Region Region within Sector School Level within Sector School Level within State (elementary and secondary public schools) Typology (private schools only) Community Type within Sector State (public schools only) School Size within Community Type within Sector Minority Status (of Students) within Community Type within Sector
School Administrator	Sector Region State (public schools only) Region within Sector School Level within Sector School Level within State (elementary and secondary public schools) Typology (private schools only)
Teacher Demand & Shortage (Private only)	Region Typology School Level Minority Status (of Students)
Teacher Demand & Shortage (Public only)	Region State Minority Status (of Students)
Teacher	Sector Region Region within Sector Minority Status (of Students) within Sector State (public schools only)

SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91.

Table 2.3 -- Definition of subpopulations in 1990-91 SASS

Subpopulation	Definition
Sector	Public or Private Schools
Region	
Northeast	Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Pennsylvania
Midwest	Ohio, Indiana, Illinois, Michigan, Wisconsin, Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, Kansas
South	Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, Oklahoma, Texas
West	Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada, Washington, Oregon, California, Alaska, Hawaii
School Level	Elementary (no grade higher than 8 and at least one of grades 1-6), Secondary (grades 7-12), and Combined (any other combination of grades; e.g., 4-9, or 5-12)
Typology	The private school typology separates private schools into three major groups and within each group into three subgroups: Catholic (parochial, diocesan, and private order), other religious (Conservative Christian, affiliated, and unaffiliated), and nonsectarian (regular, special emphasis, special education) (McMillen and Benson 1991)
School Size	Enrollment of fewer than 150 students Enrollment of 150 to 499 students Enrollment of 500 to 749 students Enrollment of 750 or more students
Community Type	Central City includes large central cities (Central cities of Standard Metropolitan Statistical Areas (SMSAs), with populations greater than or equal to 400,000 or population densities greater than or equal to 6,000 per square mile) and mid-size central cities (central cities of SMSAs, but not designated as large central cities). Urban Fringe/Large Town includes the urban fringes of large or mid-size cities (places located within SMSAs of large or mid-size central cities and defined as urban by the U.S. Bureau of the Census) and large towns (places not located within an SMSA, but that have populations greater than or equal to 25,000 and that are defined as urban by the U.S. Bureau of the Census). Rural/Small Town includes rural areas (places that have populations of less than 2,500 and that are defined as rural by the U.S. Bureau of the Census) and small towns (places not located within SMSAs, that have populations of less than 25,000, but greater than or equal to 2,500, and that are defined as urban by the U.S. Bureau of the Census).
Minority Status	Minority enrollment (sum of all racial/ethnic groups other than white) of less than 20 percent, or greater than or equal to 20 percent.
Field of Teaching	elementary general elementary special education elementary other secondary math secondary science secondary English secondary social studies secondary vocational education secondary special education secondary other

SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91.

3. Design Effect Methodology

This section presents a discussion of the *design effect* methodology. The notion of design effects and average design effects helps develop the generalized variance functions. Also, design effect provides an alternative way to obtain approximately the sampling variance estimates.

The concept of design effect was popularized by Kish (1965) in the sixties. A complex sampling design involves stratification and clustering. Stratification generally leads to a gain in efficiency over simple random sampling, but any form of clustering usually leads to a deterioration in the efficiency of the estimate due to positive intracluster correlation among the subunits in the clusters. In order to determine the total effect of any complex design on the sampling variance in comparison to the alternative simple random sampling, one calculates the ratio of these two variances associated with an estimate, namely,

$$Deff = \frac{sampling\ variance\ of\ complex\ sample}{sampling\ variance\ of\ simple\ random\ sample}.$$

This ratio is called the design effect (Deff) of the sampling design for the estimate. This ratio measures the overall efficiency of the sampling design employed and the estimation procedure utilized to develop the estimate. With a given estimation procedure, it thus provides a vehicle for comparing two competing sampling designs having the same number of sample units. This comparison presupposes that the cost of collecting the data is the same for the two competing sampling designs; i.e., the cost is determined by the number of sample units measured.

The gain in efficiency due to stratification is usually small compared to the loss in efficiency due to clustering because, for most variables, intracluster correlation is positive and not negligible. Thus, in most cases, design effect turns out to be larger than one. Accordingly, the quantity (n/Deff) can be regarded as the effective sample size for a complex design of sample size n; the effectiveness is measured relative to a simple random sampling design. Because Deff is usually larger than one, the effective sample size is usually smaller than the actual sample size.

In a survey such as SASS, where a very large number of variables are measured, it is the usual custom to calculate the design effect for a number of similar variables grouped, as the grouping described in section 2, and then calculate their average as a measure of the efficiency of the sampling design with respect to the group of variables. For example, for the Teacher Survey, the group of statistics labeled "Teacher Proportions", and the subpopulation of public schools, we calculated design effects for 23 proportion-type variables (Volume I,

Appendix I, page I-4) with a sample size of approximately 46,700 teachers. The average Deff for this group of statistics using the 23 variables was 2.8493 (Volume I, Appendix II, page II-27).

The procedure and computational formulas are presented in the following for calculating the design effects for the three most common types of estimates: totals, means, and proportions. Let x be a characteristic with sample observations x_i, $i = 1,\ldots,n$.

3.1 Design Effect for Totals

For each total-type estimate y, such as the "total number of Hispanic K-12 students," the Deff was computed in the three steps:

(1) Simple random sample variance estimate, expressed as

$$v(y)_{SRS} = \left(\sum_{i=1}^{n} w_i\right)^2 \frac{\sum_{i=1}^{n} w_i (x_i - \bar{x}_w)^2}{n \left(\sum_{i=1}^{n} w_i - 1\right)}$$

where n is the sample size (the number of respondents), w_i are the weights, and

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}.$$

The simple random sample variance estimate could be obtained using software such as SAS or SPSS. (An illustration of the SAS code for calculating the simple random sample standard error for a total, se_{SRSTOT}, is provided in Volume II - User's Manual, section 2.1.1, of this publication.)

(2) Variance estimate from complex sample, say,

$$v(y)_{COMPLEX}$$

calculated directly by the SAS WESVAR procedure (Westat 1993) using the balanced half-sample replication method.

(3) Design effect calculated as the ratio

$$Deff_{TOT} = \frac{v(y)_{COMPLEX}}{v(y)_{SRS}}.$$

3.2 Design Effect for Means

For each mean-type estimate $\bar{x}$, such as the "average number of full-time K-12 teachers in a school," the Deff was computed in the three steps:

(1) Simple random sample variance estimate, expressed as

$$v(\bar{x})_{SRS} = \frac{\sum_{i=1}^{n} w_i (x_i - \bar{x}_w)^2}{n \left(\sum_{i=1}^{n} w_i - 1\right)}$$

where

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

and n is the sample size (the number of respondent) and w_i are the weights. The simple random sample variance estimate could be obtained using software such as SAS or SPSS. (An illustration of the SAS code for calculating the simple random sample standard error for an average, se_{SRSAVG}, is provided in Volume II - User's Manual, section 2.1.1, of this report.)

(2) Variance estimate from complex sample, say,

$$v(\bar{x})_{COMPLEX}$$

calculated directly by the SAS WESVAR procedure (Westat 1993) using the balanced half-sample replication method.

(3) Design effect calculated as the ratio

$$Deff_{AVG} = \frac{v(\bar{x})_{COMPLEX}}{v(\bar{x})_{SRS}}.$$

3.3 Design Effect for Proportions

For each proportion-type estimate p, such as the "proportion of schools which have students eligible for free or reduced-price lunches," the Deff was computed in the three steps:

(1) Simple random sample variance estimate, expressed as

$$v(p)_{SRS} = \frac{p(1-p)}{n}$$

where p denotes the estimate of proportion for a characteristic of interest, expressed as

$$p = \frac{\sum_{1}^{n} w_i I(i)}{\sum_{1}^{n} w_i}$$

where $I(i)=1$ if the characteristic is present for the sampled unit and 0 if it is absent.

(2) Variance estimate from complex sample, say,

$$v(p)_{COMPLEX}$$

as calculated directly by the SAS WESVAR procedure (Westat 1993) using the balanced half-sample replication method.

(3) Design effect calculated as the ratio

$$Deff_{PROP} = \frac{v(p)_{COMPLEX}}{v(p)_{SRS}}.$$

3.4 Average Design Effects

In a large scale sample survey like SASS, data are collected for a large number of variables. Clustering does not affect all the variables in the same way as the intracluster correlation varies over all the variables. This necessitates that the Deff be computed for at least some key variables. The average of these Deffs can be considered as a measure of the efficiency of a survey sampling design compared to the standard simple random sampling.

For a suitably formed group of survey statistics, as described in section 2 of this report, the design effects may be considered similar. The average design effect over a subset of that group can provide an estimate for the common measure of design effect. This average design effect can then be used to calculate, for other variables in that group, an approximate variance estimate from the simple random sample variance estimate obtained elsewhere. This procedure gives an alternative way to obtain sampling variance estimates. (Illustrative examples for this procedure are presented in Volume I - User's Manual.) Accordingly, an average Deff was derived based on the Deffs calculated for the variables selected for each of the subpopulations (table 2.2) within the GVF groupings (table 2.1), and is listed in the Design Effect column of the corresponding GVF table (Volume I, appendix III). All those average design effects are also presented, collectively, in the Design Effects tables (Volume I, appendix II).

It will be shown, in section 4.1 of this volume, the idea of average design effect helps in the recognition of a basic form of the generalized variance functions.

4. GVF Methodology

In a survey like SASS, estimates can be computed for dozens of variables with respect to various school levels and by different sectors. The presentation of the sampling error for each of the estimates doubles the size of the report. In such large scale surveys, it is reasonable to make an attempt to cut down on the large volume of publication for sampling error estimates. It might be for the users themselves to calculate the standard errors associated with the survey estimates of interest. However, the statistical software for complex survey variance estimation, such as WESVAR (Westat 1993) and SUDAAN (Shah et al. 1992), are not widely available. The methodological motivation for developing general analytical techniques for variance estimation is from the desire to produce stable and precise sampling variance estimates. It has been found that simple mathematical relationships can be used to relate the variance or relative variance[1] of a survey estimator to the mean (expectation) of the estimator. Generalized variance functions (GVFs) are models of these mathematical relationships. The usual practice is to select a small subset of items from a larger group of survey items, calculate the estimate, variance estimate, and relative variance for the selected items by direct estimation methods, and then use these estimates as data to estimate the parameters of the GVF model. The specific GVF model is identified by estimating the parameters for several different candidate models using different fitting methodologies and then selecting the one that is the best fit using the criterion of highest R-squared value. If the *R*-squared value of the "best" model is still small, say, less than 0.5, the selected GVF may not be considered appropriate for use. In such a case, an appropriate GVF model could not be identified from the candidate models. Three different fitting procedures were examined in this study: the *ordinary least squares* (OLS), the *weighted least squares* (WLS), and the *iteratively reweighted least squares* (IRLS). Technical details of these fitting procedures will be described in section 4.2.1.

4.1 GVF Models

As introduced above, the method of generalizing variances consists of estimating the relative variance of an estimator by using a model. In this section, we present a number of possible GVF models and some intuitive theoretical justification.

---

[1] Relative variance, as introduced in the Introduction, p.2, is defined as the variance divided by the square of the mean.

Denote the estimator of a certain attribute of interest as $\hat{X}$ and let $X = E(\hat{X})$ be its mean (expectation). Then the relative variance, denoted as V^2, can be expressed as follows:

$$V^2 = \frac{var(\hat{X})}{X^2}$$

Most GVFs to be considered are based on the premise that the relative variance is a decreasing function of the magnitude of the mean X.

Here is a simple model which exhibits this property:

$$V^2 = A + B/X, \quad with\ B > 0. \qquad (Model\ 1)$$

The parameters A and B here are unknown and to be estimated. They depend upon the population, the sampling design, the estimator, and the X-attribute itself. Experience has shown that Model 1 above often provides an adequate description of the relationship between V^2 and X. In fact, the Census Bureau has used this model to develop GVFs for its Current Population Survey since 1947 (Hanson 1978); this model is also used to develop GVFs for the National Health Interview Survey (NHIS).

In an attempt to achieve an even better fit to the data than is possible with Model 1, here are alternative forms of the relative variance model to be considered (Wolter 1985):

$$V^2 = A + B/X + C/X^2, \qquad (Model\ 2)$$

$$\log(V^2) = A + B\log(X), \qquad (Model\ 3)$$

$$V^2 = (A + BX)^{-1}, \qquad (Model\ 4)$$

$$V^2 = (A + BX + CX^2)^{-1}, \qquad (Model\ 5)$$

where

V^2	=	relative variance
X	=	mean (expectation) of the selected survey estimate
A, B, C	=	unknown parameters to be estimated

In the following we give some intuitive theoretical justification for Model 1 from several aspects (Wolter 1985) to help the understanding of the GVF methodology.

(1) Suppose that the population is composed of N clusters, each of size M. A simple random sample of n clusters is selected, and each elementary unit in the selected clusters is enumerated. Then, the variance of the Horvitz-Thompson estimator $\hat{X}$ of the population total X is

$$\sigma^2 = (NM)^2 \frac{N - n}{N - 1} \frac{PQ[1 + (M-1)\rho]}{nM}$$

where $P = X/NM$ is the population mean per element and $Q = 1-P$, and ρ denotes the intraclass correlation between pairs of elements in the same cluster. The relative variance of $\hat{X}$ is

$$V^2 = \frac{N - n}{N - 1} \frac{Q}{P} \frac{[1 + (M-1)\rho]}{nM},$$

and assuming that the first stage sampling fraction is negligible, we may write

$$V^2 = \frac{1}{X} \frac{NM[1 + (M-1)\rho]}{nM} = \frac{[1 + (M-1)\rho]}{nM}.$$

Thus, for this simple sampling scheme and estimator, Model 1 provides a plausible model for relating V^2 to X. If the value of the intraclass correlation is constant (or approximately so) for a certain class of survey estimates, then Model 1 may be useful for estimating the variance of estimates in the class.

(2) If we assume an arbitrary sampling design leading to a sample of n units from a population of size N, then the design effect for $\hat{X}$ is defined by

$$Deff = \sigma^2/(N^2PQ/n),$$

where $P = X/N$ and $Q = 1-P$. This is the variance $\hat{X}$ of given the true sampling design, divided by the variance given simple random sampling. Thus, the relative variance may be expressed by

$$V^2 = Q(Pn)^{-1}Deff$$

$$= -Deff/n + (N/n)Deff/X.$$

Assuming that the Deff may be considered independent of the magnitude of X within a given class of survey statistics, the relative variance above is of the form of Model 1 and may be useful for estimating variances.

(3) Suppose it is desired to estimate the proportion $R = X/Y$, where Y is the total number of individuals in a certain subpopulation and X is the number of those individuals with a certain attribute. If $\hat{X}$ and $\hat{Y}$ denote estimators of X and Y, respectively, then the natural estimator of R is $\hat{R} = \hat{X} / \hat{Y}$. Utilizing a Taylor series approximation and assuming $\hat{Y}$ and $\hat{R}$ are uncorrelated, it can be shown (Hansen et al. 1953, Vol. II) that

$$V_R^2 = V_X^2 - V_Y^2,$$

where V^2_R, V^2_X, and V^2_Y denote the relative variances of $\hat{R}$, $\hat{X}$, and $\hat{Y}$ respectively. If Model 1 holds for both V^2_X and V^2_Y, then V^2_R above gives

$$V_R^2 \doteq \frac{B}{X} - \frac{B}{Y}$$

$$= \frac{B}{Y}\frac{(1 - R)}{R},$$

and hence

$$var(\hat{R}) \doteq \frac{B}{Y} R(1 - R).$$

The above equation for var($\hat{R}$) has the important property that the variance of an estimator

$$\hat{R}_1 = \hat{X}_1 / \hat{Y}$$

of a proportion

$$R_1 = X_1 / Y$$

which satisfies

$$R_1 = 1 - R$$

is identical to the variance of the estimator $\hat{R}$ of R. Thus, for example, var($\hat{R}$) = var(1- $\hat{R}$). Tomlin (1974) justifies Model 1 on the basis that it is the only known model that possesses this important property.

Attempts have been made to develop the theory to justify the use of GVFs, in particular, the Model 1. Valliant (1987) established, for Model 1, asymptotic theory for estimators of totals that are linear combinations of sample cluster means from stratified two-stage cluster samples. The validness of Model 1 has also been recognized by many empirical studies including the above Valliant (1987), especially, for binary variables.

4.2 GVF Estimation for 1990-91 SASS

NCES recently conducted a study to determine the feasibility of including generalized variance functions in SASS publications (Synectics 1992). That study provided a thorough examination of five different GVF models using three different fitting procedures (see section 4.2.1) for the 1987-88 SASS. Preliminary analysis determined only three of the five models were viable GVFs for the groups of estimates under study. The determination for the best GVF of the remaining three models was based on comparing three different fitting methodologies. The final choice of GVFs from that earlier effort may no longer be applicable to the 1990-91 SASS, due to some significant changes in the sample allocation. However, because the three models evaluated for the 1987-88 include the most recognized GVFs by various empirical studies, the current work adopts the earlier conclusion that only those three models would be viable GVFs.

4.2.1 Candidate Models and Fitting Methodologies

As a result of the 1987-88 SASS generalized variance estimation effort (Synectics 1992), only Models 1, 3, and 4 of the five models described in section 4.1 were determined to be viable candidates for estimating GVF parameters for the 1990-91 SASS. For computational reasons, the actual models used in the fitting were the coefficient of variation[2] (CV) version of these models, that is, $CV = (A + B/X)^{1/2}$ for Model 1, $\log(CV) = A + B \log(X)$ for Model 3, and $CV = (A + BX)^{-1/2}$ for Model 4.

The following three different fitting procedures were also examined to determine the "best" model fitting technique: the ordinary least squares (OLS), the weighted least squares (WLS), and the iteratively reweighted least squares (IRLS). (Note that the weights used in the WLS and IRLS procedures do not refer to the sample weights.) The OLS procedure was specified to work with the unweighted sum of residual squares. The WLS procedure was, in particular, specified to work with the sum of residual squares which weight inversely to the square of the observed CV, and the IRLS method was specified to work with the sum of residual squares which weight inversely to the square of the predicted CV with the weights updated at each iteration. Based on our investigation, the WLS technique was determined to be the best. As is known, the OLS technique gives too much weight to the small estimates whose corresponding relative variances are usually large and unstable. The WLS technique is better than the OLS because it gives a reduced weight to the least reliable terms in the sum of residual squares. The IRLS technique has the same advantage as WLS but may give somewhat different results.

---

[2] Coefficient of variation is the square root of the relative variance.

During the course of the current effort using 1990-91 SASS data, it was concluded that Model 4 was no longer an appropriate model due to its failure to produce parameter estimates because of the lack of convergence of the iterative fitting procedure for many of the groups of estimates. Furthermore, when it did converge, the Model 4 fit often resulted in a possible negative bias (understatement) of CV for large estimates.

4.2.2 GVF Procedure

The basic GVF procedure used for variance estimation for each of the 16 groups of statistics (table 2.1) and for each of the relevant subpopulations (table 2.2) is summarized in the following steps. Results and conclusions will be discussed in the next section.

Step 1: Grouping items prior to model estimation

Building on the final set of variables used in model estimation during the 1987-88 GVF effort (Synectics 1992), a provisional set of variables (on average approximately 25 for each type of group) was selected (see appendix I). Estimates of totals, averages, and proportions for these selected variables were calculated. This was followed by a direct calculation of the relative variance and coefficient of variation (CV) of each of these statistics, using a balanced half-sample replication technique. These estimates were chosen as a provisional group of similar items to be used for model estimation. Final groups of statistics for model estimation were determined by examining design effects and simply removing from the provisional set those statistics that appeared to follow a different model than the majority of the statistics in the group. Other statistics, originally outside the provisional set, but appeared consonant with the group model, were then added. Scatter plots of the logarithm of the CV versus the logarithm of the estimate were examined to form the "final" groups of statistics that would follow a common model. The success of the GVF technique depends critically on the grouping of the survey statistics. However, only a limited number of choices was available for the SASS surveys from which variables were chosen for each of the 16 groups of statistics.

Step 2: Estimating model parameters

Using the final group of statistics and their respective CVs calculated in Step 1, Models 1 and 3 were fitted through the different fitting procedures described in section 4.2.1, using the statistical package SAS nonlinear regression procedure, NLIN. Specifications for the NLIN procedure included requesting estimates of the parameters and the respective R-squared values. The iterative method specified for the NLIN procedure

was the modified Gauss-Newton method which regresses the residuals onto the partial derivatives of the model with respect to the parameters until the estimates converge. The final parameter values from the earlier 1987-88 GVF work, as useful information from a previous study, were used as starting values in the current iterative runs.

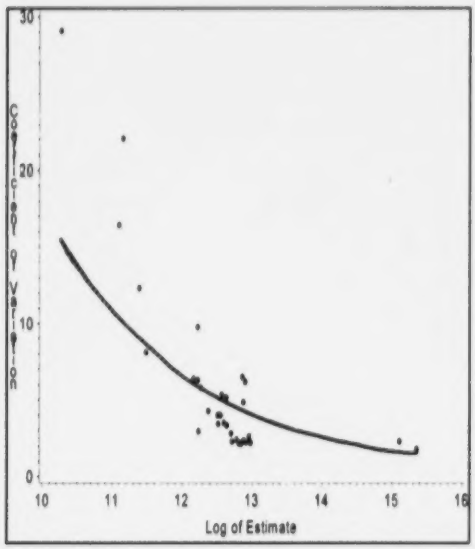
Note: Results from the IRLS procedure were not promising in terms of convergence, and the weighted least square procedure was judged the most appropriate. Table 5.1 shows the advantage of WLS over IRLS in this study.

Step 3: Determining best model fit

In determining the best model fit, it was useful to examine, for each of the models, the overlay plot of the fitted CV regression curve onto the scatter plot of the CV data, using the logarithm of the estimate as the reference (for graphical presentation reasons). Such a plot is a kind of predicted-versus-observed plot. How well the shape of the curve accords with the observed reality gives a visual exhibition of the goodness of fit. Illustrating this type of graphic presentation, figures 4.1 - 4.6 show the fitted Model 1 curve overlaying the scatter plot of the coefficient of variation for various statistics in the following subpopulations: student totals, school proportions, teacher totals, teacher proportions, administrator totals, and administrator proportions. Each figure contains two plots: (a) for weighted fitting and (b) for iteratively reweighted fitting. A comparison may be made between the two fitting procedures.

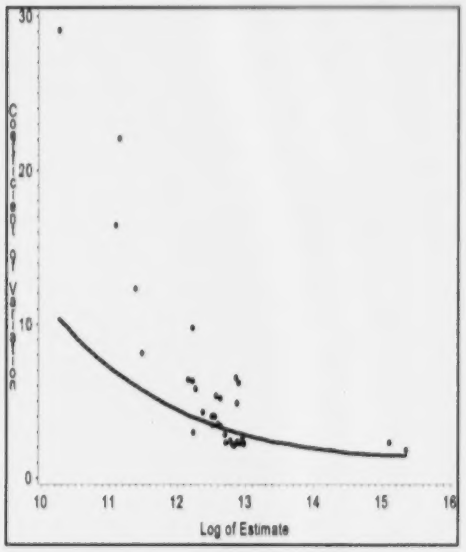
However, when evaluating the fit of the models, a widely used single index is the *R-squared value* which is defined as one minus the ratio of the sum of residual squares from the model divided by the total sum of squares of the dependent variable. This measure can be interpreted as the percentage of the variation of the dependent variable being explained by the model. Thus, an R-squared value close to one shows a good fit. We calculated the R-squared values to compare the two candidate models each with two fitting procedures. R-squared is a good measure of fit because the models we were evaluating each had two parameters. One should not compare models with different numbers of parameters based on the R-squared alone.

Figure 4.1(a) -- Weighted fitted line and scatter plot for groups: School Survey / school totals (students) / private



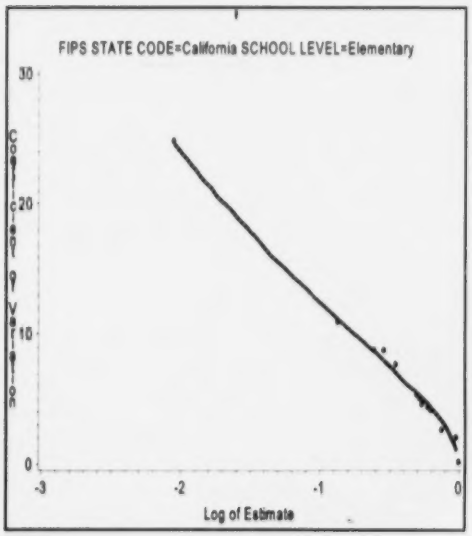
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91 (School Questionnaire)

Figure 4.1(b) -- Iteratively reweighted fitted line and scatter plot for group: School Survey / school totals (students) / private



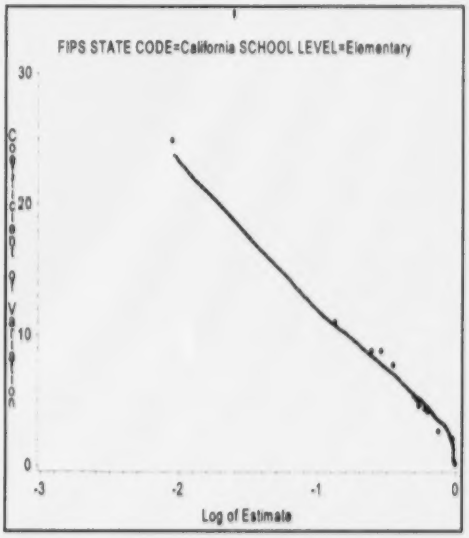
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91 (School Questionnaire)

Figure 4.2 (a) -- Weighted fitted line and scatter plot for groups: School Survey / school proportions / California / elementary



SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91 (School Questionnaire)

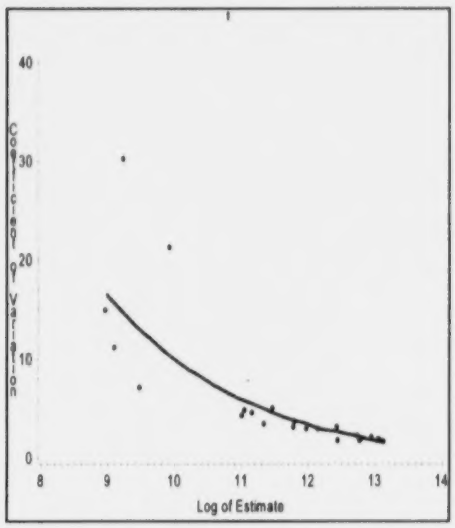
Figure 4.2 (b) -- Iteratively reweighted fitted line and scatter plot for group: School Survey / school proportions / California / elementary



SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91 (School Questionnaire)

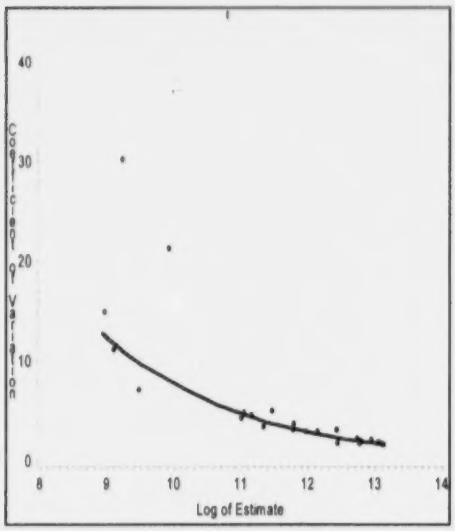
BEST COPY AVAILABLE

Figure 4.3(a) -- Weighted fitted line and scatter plot for group: Teacher Survey / teacher totals / west



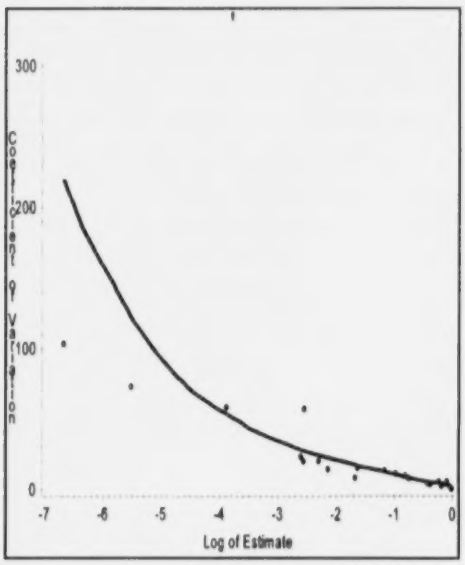
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91 (Teacher Questionnaire)

Figure 4.3(b) -- Iteratively reweighted fitted line and scatter plot for group: Teacher Survey / teacher totals / west



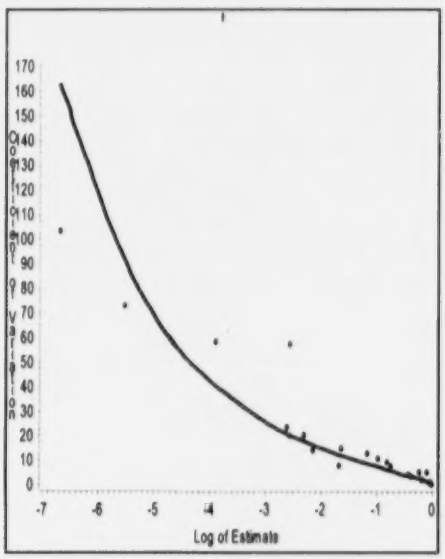
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91 (Teacher Questionnaire)

Figure 4.4 (a) -- Weighted fitted line and scatter plot for groups: Teacher Survey / teacher proportions / Vermont



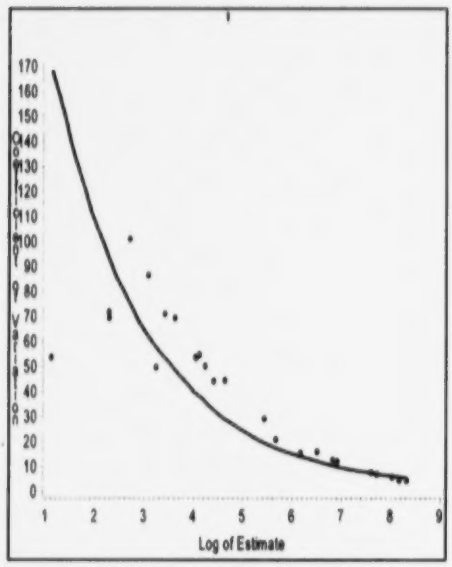
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91 (Teacher Questionnaire)

Figure 4.4 (b) -- Iteratively reweighted fitted line and scatter plot for group: Teacher Survey / teacher proportions / Vermont



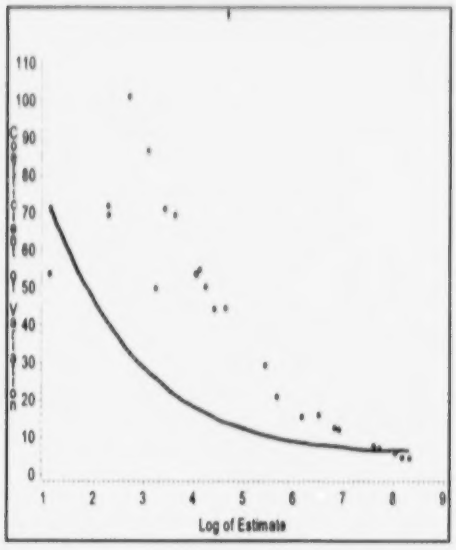
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91 (Teacher Questionnaire)

Figure 4.5(a) -- Weighted fitted line and scatter plot for group: School Administrator Survey / administrator totals / Catholic Parochial



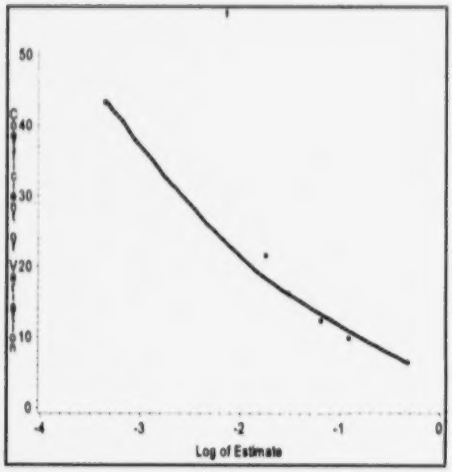
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91 (School Administrator Questionnaire)

Figure 4.5(b) -- Iteratively reweighted fitted line and scatter plot for group: School Administrator Survey / administrator totals / Catholic Parochial



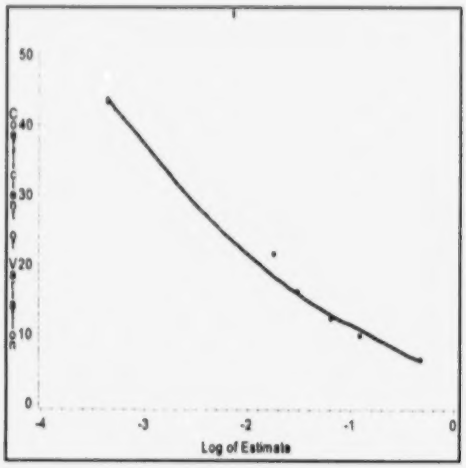
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91 (School Administrator Questionnaire)

Figure 4.6 (a) -- Weighted fitted line and scatter plot for groups: School Administrator Survey / administrator proportions / West Virginia



SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91 (School Administrator Questionnaire)

Figure 4.6 (b) -- Iteratively reweighted fitted line and scatter plot for group: School Administrator Survey / administrator proportions / West Virginia



SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91 (School Administrator Questionnaire)

BEST COPY AVAILABLE

Step 4: Reducing the number of distinct GVF tables

In the interest of reducing the number of distinct GVF tables, we conducted an empirical investigation to determine if a simplified approach could be applied to estimate the variance of the averages using the GVFs developed for the totals. The simplified approach is to use the following formula to derive approximately the standard error of an **average** from the standard error of the corresponding **total**:

$$se_{AVG} = \frac{se_{TOT}}{\sum_{1}^{n} w_i}$$

where, on the right-hand side, se_{TOT} is the standard error of a total type estimate calculated directly (e.g., by the balanced half-sample replication method) or through the GVF, and w_i are the weights. The above formula is approximate because the domain over which the weights are summed (in the denominator) can vary randomly.

The empirical investigation compared, for a number of survey items, the standard errors for average obtained by the two approaches: directly estimated by the BRR method, and derived by the above formula from the standard error of the corresponding total which was estimated directly by the BRR method. The results of the comparison regarding which GVF tables could be reduced are discussed in section 5.4 of this volume.

More details for the use of the above formula are provided in section 3.3, Volume I - User's Manual, of this publication.

5. Results and Conclusions

5.1 Average Design Effects

Appendix II includes the tables of average design effects (see sections 3.1, 3.2, and 3.3) calculated for each of the groups of survey estimates of interest: totals, averages, and proportions, based on the subpopulations for each of the 1990-91 SASS surveys.

In calculating the average design effects for each of the subpopulations identified in table 2.2, some unusually high design effects occurred. A few individual variables were identified with very high design effects, high enough to raise the average design effect for the subpopulation to a questionable value. These highly skewed design effect values were correlated with a small number of observations for that variable in the subpopulation of interest. Removal of the design effect for these variables from the calculation of the average design effect would produce more homogeneous average design effects. For example, in the School Survey, for subpopulations corresponding to school size within community type within sector, we removed all records with a design effect of greater than 30. Eighteen observations (or approximately two percent) were removed by this condition, and the resulting drop in the average design effect brought the numbers into line with other cuts. More specifically, for school size of 500 to 749 in central city communities within the public school sector, removal of the design effect corresponding to the variable TOTALENR (number of students enrolled in K-12 grades plus ungraded as of October 1 of this school year) from the average design effect calculation caused the average design effect to drop from 10.6512 to 7.3773. A similar pattern of highly skewed design effects corresponding to small sample sizes for particular variables occurred in multiple subpopulations across other survey components. We deleted these variables from the average design effect calculation in these cases. There were also cases where very high design effects corresponded to variables that were counting almost everyone. Here the sample sizes were not small. An example of such a variable was ASC017 - Have a Master's Degree for the Administrator Survey. These skewed variables were also deleted from the average design effect calculations. Table 5.1 below provides a listing of all variables deleted from the average design effect calculations, their individual design effect value, and the corresponding re-calculated new average design effect for the particular subpopulation after removing the particular variable(s). However, for the School Administrator Survey, because not many variables were included in the calculation of average design effects, highly skewed variables were not removed from the calculations.

There are a very large number of certainty and high probability districts in the public Teacher Demand and Shortage Survey (TDS) sample. These districts also contain a very large proportion of the total number of teachers and students. For the complex SASS design, these districts contribute very little to the variance estimates of totals and averages. However, for a

simple random sample design, these same districts contribute a very large part of the variance estimates of totals and averages. Due to these differences in variance contribution, and depending on the subpopulation, the design effects can vary greatly. Often these design effects can be extremely small (design effects less than 0.2 are not uncommon). Hence, since these findings are not realistic, using an average design effect on all variables would be inaccurate. Similarly, TDS proportions have low average design effect results, but to a lesser extent. For this reason, neither average design effects nor GVF tables are presented for the public TDS.

Table 5.1 -- Effect of Highly Skewed Design Effects on Average Design Effect Calculations

Survey Component	Groups of statistics	Subpopulation	Variable[1]	Sample Size	Deff	New Average Deff
School	Student Totals	Arkansas Elementary	NUMBRPK	3	89.5	
			NUMBR8	5	22.5	
			NUMBR7	7	9.2	1.3602412
		Kentucky Elementary	BILNGNUM	2	82.5	1.2531488
		West Virginia Secondary	BILNGNUM	2	214.3	
			AFTERNUM	2	8.7	1.5214717
Admin	Admin Totals	West Virginia Public	ASC017	164	81.6	1.1389081
		Michigan Elementary	ASC017	92	122.2	1.3771336

[1]Labels for variables are provided in Volume I, Appendix I

The following sections 5.2 - 5.4 present the results of the GVF procedure described in section 4.2. These results are compared in order to decide on the best GVF model to use across all the surveys and the validation of this model is also presented. In addition, the results of comparing the approximation formula for calculating standard errors for averages using the GVFs for totals to the direct estimate of the standard error is presented.

5.2 "Best" GVF Models

For each group of statistics, the R-squared values were compared for Model 1 WLS, Model 1 IRLS, Model 3 WLS, and Model 3 IRLS across all subpopulations. The result of this comparison is represented by a count of how many times that particular model fitting methodology produced the highest R-squared value. These counts are displayed at the group of statistics level within each survey component for each of the four combinations of the two models and the two fitting methodologies and are displayed in the columns labeled: Model 1 WLS, Model 1 IRLS, Model 3 WLS, Model 3 IRLS in table 5.2. For example, for the Teacher Survey-Teacher Averages, Model 1 WLS fit best 13 times out of 69 subpopulation comparisons. The last two rows in the table summarize the results across all survey components. The results show that overall WLS fits better than IRLS for both Model 1 and Model 3. Also, overall Model 1 WLS fits better than Model 3 WLS. We examined those cases where Model 3 WLS fit better than Model 1 WLS (e.g., student totals in the School Survey) and found the R-squared values were very close. On the other hand, in many cases, when Model 1 WLS fit better than Model 3 WLS, the R-squared values were significantly better. In conclusion, as the last row in the table shows, Model 1 with weighted least squares fitting provides the best GVF overall. Thus, the GVF tables provided in Volume I, Appendix III, only include the parameters resulting from the Model 1 WLS fitting procedure.

Table 5.2 -- Comparison of models and fitting techniques for estimating standard errors

Survey component	Groups of statistics	Model 1 WLS	Model 1 IRLS	Model 3 WLS	Model 3 IRLS	Number of comparisons
Teacher	Teacher avgs.	13	0	56	0	69
	Teacher totals	29	3	35	2	69
	Teacher prop.	47	15	6	0	69
Teacher Demand and Shortage	Averages	19	0	0	0	19
	Proportions	14	3	2	0	19
	Totals	19	0	0	0	19
School	Student avgs.	58	17	84	42	224
	Student totals	85	20	118	0	224
	Teacher avgs.	115	1	104	0	220
	Teacher totals	143	3	42	35	224
	School prop.	102	79	6	1	188
Administrator	Admin. avgs.	104	0	77	0	182
	Admin. totals	74	79	28	0	182
	Admin. prop.	82	86	14	0	182
Total Times a Model Fits Best		904	306	572	80	1890
Percent of Times a Model Fits Best		47.8%	16.2%	30.3%	4.2%	--

Model 1 corresponds to the following: $CV = (A + B/X)^{1/2}$
Model 3 corresponds to the following: $\log(CV) = A + B \log(X)$
WLS stands for weighted least squares.
IRLS stands for iteratively reweighted least squares.
Avgs. stands for averages.
Prop. stands for proportions.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91.

5.3 GVF Validation

For each survey, the final fitted Model 1 GVFs were applied to selected variables that had been held in reserve for validation purposes. Table 5.3 presents the results of using these variables to compute a "relative error of prediction":

$$\text{relative error} = (se_{WESVAR} - se_{GVF}) / se_{WESVAR}$$

which thus expresses the under- or overestimate of the GVF standard error as a proportion of the standard error directly estimated using a balanced half-sample replication method (by WESVAR). Columns 1 and 2 identify the survey component/group of statistics and the name/label of the variable used in the validation. The third column provides the measure of fit (R-squared value) of the GVF model for each of the survey components/group of statistics provided in table 5.3. The fourth column presents the result of calculating the standard error directly using PROC WESVAR and the fifth column presents the result of calculating the corresponding standard error using the GVF. Finally, the last column presents the value of the percent relative error of prediction as defined above.

Table 5.3 -- Comparison of directly estimated standard error vs. GVF standard error for selected variables

Survey / Type of estimate	Selected variable	GVF *R*-squared	se directly estimated	se by GVF	Relative error (%)
Teacher Survey / Totals (Northeast)	MTSC022: Classif. of main activity prior year to	0.9401	2807.51	2827.8011	0.72
Teacher Survey / Totals (Midwest)	MTNEWID: New teacher indicator	0.9685	2199.22	3157.9457	43.59
Teacher Survey / Totals (South)	MTSC022: Classif. of main activity prior year to	0.9467	3032.69	3061.877	0.96
Teacher Survey / Totals (South)	MTSC048: Any other degrees	0.9467	3752.83	3804.4736	1.38
Teacher Survey / Proportions (Public/Less than 20%)	PTSC290: Work a nonschool job summr '90-end '91?	0.9755	0.00399	0.004008264	0.46
Teacher Survey / Proportions (Private/Less 20%)	PTSC286: Summer school earnings summr '90 (y/n)	0.9923	0.00705	0.00707319	0.33
Teacher Survey / Proportions (Private/20% or greater)	PTSC290: Work a nonschool job summr '90-end '91?	0.9613	0.00913	0.008994817	-1.48
School Administrator Survey / Totals (Public/Northeast)	ASCO12: Have a bachelors degree	0.9139	66.89647	65.11464322	-2.66
School Administrator Survey / Totals (Public/South)	ASC024: Have earned an ed spec/prof dip.	0.9747	232.9416	242.4122285	4.07
School Administrator Survey / Totals (Public/South)	ASC028: Mjr. field of study for doctorate	0.9747	106.2261	107.0300084	0.76
School Administrator Survey / Totals (Private/Northeast)	ASC028: Mjr. field of study for doctorate	0.8807	30.3029	25.48260047	-15.91
School Administrator Survey / Totals (Private/Midwest)	ASC027: Earned a doctorate/1st prof deg.	0.9181	67.62073	64.11382096	-5.19
School Administrator Survey / Totals (Private/Midwest)	ASC028: Mjr. field of study for doctorate	0.9181	41.41898	43.41752102	4.83
School Administrator Survey / Totals (Private / West)	ASC024: Have earned an ed spec/prof dip.	0.8653	60.0311	47.44266961	-20.97

Table 5.3 --Comparison of directly estimated standard error vs. GVF standard error for selected variables (cont.)					
Survey / Type of estimate	**Selected variable**	**GVF *R*-squared**	**se directly estimated**	**se by GVF**	**Relative error (%)**
School Survey / Student Totals (Public/Rural-Small Town/ 150-499)	Total enrollment	0.8803	147998.7	135567.0833	-8.40
School Survey / Student Totals (Private/Central City/1-149)	Number days in school year	0.6061	44471.02	57169.53123	28.55
School Survey / Student Totals (Private/Urban Fringe-Large Town/500-749)	Total enrollment	0.7970	32230.06	35135.53576	9.01
School Survey / Student Totals (Private/Rural-Small Town/750)	Number days in school year	0.1644	1588.194	1631.578548	2.73
School Administrator Survey / Proportions (Elementary/ Mississippi)	PASC121: Are you male or female?	0.9583	0.05689	0.05728118	0.69
School Administrator Survey / Proportions (Elementary/New York)	PASC082: Level of teachers' verbal abuse	0.8875	0.05797	0.043796359	-24.45
School Administrator Survey / Proportions (Elementary/Virginia)	PASC086: Poverty level	0.9019	0.0369	0.035349632	-4.20
School Administrator Survey / Proportions (Elementary/Alaska)	PASC020: Do you have any other type of degree?	0.9948	0.05509	0.054878061	-0.38
School Administrator Survey / Proportions (Elementary/Idaho)	PASC082: Level of teachers' verbal abuse	0.4836	0.05079	0.051247441	0.90
School Administrator Survey / Proportions (Elementary/Maine)	PASC085: Level of parental alcoholism and/or drug	0.9925	0.03808	0.037798761	-0.74
School Administrator Survey / Proportions (Secondary/Arizona)	PASC086: Poverty level	0.9701	0.04262	0.043169401	1.29
School Administrator Survey / Proportions (Secondary/California)	PASC087: Level of racial tension	0.9643	0.04002	0.040605839	1.46
School Administrator Survey / Proportions (Secondary/D.C.)	PASC121: Are you male or female?	0.932	0.12728	0.126846388	-0.34
School Administrator Survey / Proportions (Secondary/Massachusetts)	PASC087: Level of racial tension	0.9602	0.02648	0.026027905	-1.71

Table 5.3 --Comparison of directly estimated standard error vs. GVF standard error for selected variables (cont.)					
Survey / Type of estimate	**Selected variable**	**GVF *R*-squared**	**se directly estimated**	**se by GVF**	**Relative error (%)**
School Administrator Survey / Proportions (Secondary/North Carolina)	PASC084: Level of lack of parent involvement	0.8913	0.05064	0.053759434	6.16
School Administrator Survey / Proportions (Secondary/ Oklahoma)	PASC020: Do you have any other type of degree?	0.9792	0.04907	0.046852214	-4.52
School Administrator Survey / Totals (Public)	ASC015: Have a 2nd Mjr. or Minor field of study	0.9533	501.3934	481.3949434	-3.99
School Administrator Survey / Totals (Private)	ASCO12: Have a bachelors degree	0.9284	175.4621	177.02689	0.89
School Survey / Student Totals (Public/Central City)	GRADAPLY: Nmbr last yr grads applied to 2/4 colleges	0.6182	25019.01	22532.62945	-9.94
School Survey / Student Totals (Private/Central City)	GRADAPLY: Nmbr last yr grads applied to 2/4 colleges	0.8751	6738.889	6802.797814	0.95
School Survey / Student Totals (Private/Urban Fringe-Large Town)	GRADAPLY: Nmbr last yr grads applied to 2/4 colleges	0.7697	6522.346	5393.470758	-17.31
School Survey / School Proportions (Northeast)	ELSENUM: Nmbr stdnts attend othr sch part of day	0.9192	0.0039015	0.004769906	22.26
School Survey / School Proportions (South)	ELSENUM: Nmbr stdnts attend othr sch part of day	0.9628	0.0025075	0.003425276	36.60
School Survey / School Proportions (West)	ELSENUM: Nmbr stdnts attent othr sch part of day	0.8841	0.0038424	0.004428094	15.24

5.4 Reduction of GVF Tables

The results of the empirical investigation to determine if a simplified approach of using the GVFs developed for the total to estimate the standard error of the averages are presented in this section. The empirical investigation used a set of variables in the School Survey and the Teacher Survey to compare the standard errors for averages using the two approaches - directly estimated vs. derived - as described in Step 4 of section 4.2.2. Table 5.4 gives the comparison. The first two columns identify the survey component/subpopulation and the variable, respectively. The third column gives the sum of weights for that variable. Column 4 presents the standard error directly estimated by the BRR method, and column 5 presents the derived standard error using the approximation formula (see section 4.2.2, Step 4). Finally, column 6 provides the relative percent difference between column 4 and 5 which expresses the under- or overestimate of the derived standard error as a proportion compared to the directly estimated standard error. Table 5.5 lists the names and labels of the variables used in table 5.4.

The results in table 5.4 show that for variables from the School Survey, the derived standard error appears close to the directly estimated standard error. Thus, for the School Survey, it seems reasonable to reduce the GVFs for the averages from the set of GVF tables. The GVFs for the totals can be used to derive the standard errors for the averages. For an illustrative example, see volume I, section 3.3.

On the other hand the simplified approach does not do as well for the variables in the Teacher Survey. Most of the variables are counting the number of courses taken or time spent. For these types of variables, when the average is of interest, the corresponding total does not seem of the same degree of interest. (For example, the average number of courses a public school teacher taught in the state of Arizona would be of interest, while the total number of courses the teachers of the state taught is, though making sense as a measure of the size of education for the state, not widely used.) Also, table 5.4 shows for the variables from the Teacher Survey the difference is quite big between the formally derived standard error and the directly estimated standard error. Therefore, it is necessary to include the GVF table for teacher averages from the Teacher Survey in the set of final GVF tables.

Table 5.4 -- Comparison of the directly estimated and derived standard errors

Survey & subpopulation	Variable	Sum of weights	Directly estimated se	Derived se (by formula)	Relative difference[1] (%)
School Public/Central City	HISPNSTU	18683.82	5.3435	5.472	2.4
	NUMBR6	8578.40	4.0654	5.3213	31
School Public/Rural-Small Town	GRADNUM	11653.22	2.18127	2.4747	13
	MATHNUM	24783.24	1.04758	1.3097	25
School Private/Catholic Parochial	BLACKSTU	5436.85	3.4544	3.6382	5.3
	NUMBR6	4928.18	1.1930	1.5549	30
School Private/Nonsectarian Regular	NUMBR12	797.58	3.1319	3.1830	1.6
	GRADNUM	785.66	3.0473	3.6854	21
School Public/California	FULTEACH	1074.35	0.78948	0.78706	-0.3
	ASIANTCH	1074.35	0.02022	0.02020	-0.1
	ELEMNEW	957.31	0.13634	0.14052	3.0
	ENGLNEW	957.31	0.03909	0.03823	-2.2
	LFTTEACH	810.37	0.10003	0.11105	11
	ABSNTCH	1074.35	0.09486	0.09768	3.0
Teacher Public/Minority LE20%	TSC078	2981.17	0.098	0.4229	331
	TSC079	5504.80	0.115	0.2686	134
	TSC082	10057.34	0.077	0.1386	80

[1] Relative difference (in percent) = 100 x (derived *se* - directly estimated *se*) / (directly estimated *se*).

Table 5.5 -- Variables used in table 5.4 for comparison of standard errors

SCHOOL SURVEY	
ABSNTCH	Number K-12 teachers absent most recent day
ASIANTCH	Number asian/pacific islander K-12 teachers
BLACKSTU	Number black/non-hispanic K-12 student
ELEMNEW	Number new K-12 teachers main assignment: elmentary
ENGLNEW	Number new K-12 teachers main assignment: English
FULTEACH	Number full-time K-12 teachers
GRADNUM	Number 12th grade students graduated last year
HISPNSTU	Number hispanic K-12 students
LFTTEACH	Number K-12 teachers left teaching
MATHNUM	Number remedial mathematics students
NUMBR6	Number students enrolled in 6th grade
NUMBR12	Number students enrolled in 12th grade
TEACHER SURVEY	
TSC078	Number undergraduate math courses taken
TSC079	Number graduate math courses taken
TSC082	Number graduate computer science courses taken

SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey: 1990-91

The final GVF tables, as the products of this study, are presented in Volume I, the User's Manual, Appendix III, of this publication. The following is a list of the final GVF tables provided:

The School Survey:

- GVFs for Student Totals
- GVFs for Teacher Totals
- GVFs for School Proportions

The School Administrator Survey:

- GVFs for Administrator Totals
- GVFs for Administrator Proportions

The TDS Survey:

- GVFs for Totals (Private Schools)
- GVFs for Proportions (Private Schools)

The Teacher Survey:

- GVFs for Teacher Totals
- GVFs for Average Number of Courses Taken or Time Spent
- GVFs for Teacher Proportions

5.5 User's Manual

Volume I - User's Manual of this publication illustrates how to use the design effects and GVFs to approximate standard errors for the 1990-91 Schools and Staffing Survey (SASS). Appendix I of the manual provides a list of the variables used in developing the average design effects and GVFs. Average Design effect and GVF tables are included in appendices II and III, respectively, of the manual. Appendix IV of the manual is a sum of weights table which is used together with the GVF tables for totals to derive the standard errors for averages.

6. Next Steps

This study has focused on the production of design effects and GVFs for 1990-91 SASS users. Based on this study, we consider further steps may be taken to investigate and possibly improve on alternatives for calculating standard errors for NCES surveys. The following summarizes possible next steps.

- Examine residuals from existing GVF models and regression diagnostics to attempt to improve model fits. It is possible that formal clustering methods rather than the current subjective groupings would yield some improvement.

- Consider local fitting (Loess) type methods rather than just global models. Connected to this would be the greater "seeing power" of graphical visualization techniques applied to GVF fitting (see Cleveland 1985).

- Look at mixtures of direct variance estimates and GVFs when no good models can be found or when it is desired to guard against model failure.

- Consider applications of extreme value approaches and other techniques for highly skewed data (such as median rather than mean regression) and other robust estimators (see Hoaglin et al. 1983, 1985, and 1991).

- Carry out simulation studies of the coverage properties of estimators. Perhaps use some well-defined closeness to 95 percent nominal coverage as a measure of goodness of fit rather than using R-squared.

- Choose variables more systematically over the range of X to improve estimation - that is, employ experimental design ideas.

References

Anderson, T. W. (1958), *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley.

Box, G., Hunter, W. and Hunter, J. S. (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. New York: John Wiley.

Cleveland, W. (1985), *The Elements of Graphing Data*. Monterey, CA: Wadsworth Advanced Books and Software.

Copas, J. B. (1983), "Regression, Prediction and Shrinkage," *Journal of the Royal Statistical Society*, Ser. B, 45, 311-354.

Edelman, M. W. (1967), *Curve Fitting of Keyfitz Variances*. Unpublished memorandum. Washington, DC: U.S. Bureau of the Census.

Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953), *Sample Survey Methods and Theory. Vol. I: Methods and Applications. Vol. II: Theory*. New York: John Wiley.

Hanson, R. H. (1978), *The Current Population Survey: Design and Methodology*. Technical Paper 40. Washington, DC: U.S. Bureau of the Census.

Hoaglin, D., Mosteller, F. and Tukey, J. (1983), *Understanding Robust and Exploratory Data Analysis*. New York: John Wiley.

----------. (1985), *Exploring Data Tables, Trends, and Shapes*. New York: John Wiley.

----------. (1991), *Fundamentals of Exploratory Analysis of Variance*. New York: John Wiley.

Johnson, E. G. and King, B. F. (1987), "Generalized Variance Functions for a Complex Sample Survey," *Journal of Official Statistics*, 3, 235-250.

Kaufman, S. and Huang, H. (1993), *1991 Schools and Staffing Survey: Sample Design and Estimation*. Technical Report NCES 93-449. Washington, DC: National Center for Education Statistics.

Kish, L. (1965), *Survey Sampling*. New York: John Wiley

Lee, K. H. (1972), "Partially balanced designs for half sample replication method of variance estimation," *Journal of the American Statistical Association* 67, 324-334.

McCarthy, P. J. (1966), "Replication: An Approach to the Analysis of Data from Complex Surveys," *Vital and Health Statistics*, Series 2, No. 14, U.S. Department of Health, Education and Welfare. Washington, DC: U.S. Government Printing Office.

McMillen, M. and Benson, P. (1991), "Diversity in Private Schools." Technical Report NCES 92-082. Washington, DC: U.S. Department of Education.

Rothhaas, R. (1993), *CPS New and Improved Standard Error Parameters for Labor Force Characteristics*. Unpublished memorandum. Washington, DC: U.S. Bureau of the Census.

Rust, K. (1986), "Efficient Replicated Variance Estimation," *Proceedings of the American Statistical Association Section on Survey Research Methods*, 81-87.

Shah, B.V., Barnwell, B.G., Hunt, P., and LaVange, S.C. (1992), *SUDAAN User's Guide* (software version 6.00, 1992). Research Triangle Park, NC: Research Triangle Institute.

Short, K. S. and Littman, Mark S. (1989), *Transitions in Income and Poverty Status: 1984-85*. Current Population Reports, Series P-70, No. 15-RD-1. U.S. Government Printing Office. Washington, DC: U.S. Bureau of the Census.

Synectics for Management Decisions, Inc. (1992), *Generalized Variance Estimates for 1987-88 Schools and Staffing Survey (SASS)*. NCES Working Paper Series 94-02. Washington, DC: National Center for Education Statistics.

Tomlin, P. (1974), *Justification of the Functional Form of the GATT Curve and Uniqueness of Parameters for the Numerator and Denominator of Proportions*. Unpublished memorandum. Washington, DC: U.S. Bureau of the Census.

Valliant, R. (1987), "Generalized Variance Functions in Stratified Two-Stage Sampling," *Journal of the American Statistical Association*, 82, 499-508.

Westat, Inc. (1993), *The WESVAR SAS Procedure Version 1.2*. Rockville, MD: Westat, Inc.

Wolter, K. M. (1985), *Introduction to Variance Estimation*. New York: Springer-Verlag.

ISBN 0-16-048504-5

90000

52 9 780160 485046

United States
Department of Education
Washington, DC 20208–5651

Official Business
Penalty for Private Use, $300